

Remarks

The Examiner has rejected all pending claims under 35 U.S.C. 103(a) over Xactware (www.xactware.com) combined with Bunte (U.S. 6,330,975), Chainer (U.S. 6,397,334), Bradbury (U.S. 5,442,512) and/or Harvey (U.S. 6,208,507).

In paragraph 5, the Examiner states that Xactware was “retrieved from the Internet Archive Wayback Machine <www.archive.com>, 6/29/1998.” The Applicants respectfully traverse the Examiner’s rejections based on Xactware because the Internet Archive was not publically available until October 2001. (Exhibit A, p. 1, “With the October 2001 launch of the Wayback Machine, this huge archive is now freely available to the Web public.”)

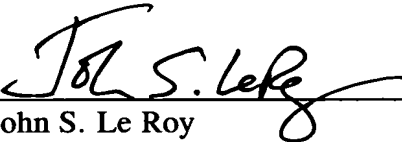
To be a “printed publication,” a reference must be publically disseminated and/or publically accessible *i.e.* archived or indexed in such a manner that an interested member of the public can locate the reference. *In re Klopfenstein*, 380 F.3d 1345, 1348 (Fed.Cir. 2004). For example, a thesis residing in a public library that is not catalogued in a searchable fashion is not a printed publication, regardless of the fact that the thesis physically resided in the public domain. *In re Cronyn*, 890 F.2d 1158, 1161 (Fed.Cir.1989).

World Wide Web postings stored in the Internet Archive were not “publically accessible” by way of the Wayback Machine prior to the applicants’ filing date, April 12, 2000. The Examiner has made no showing that an interested member of the public could have searched a publically-available database for the posting prior to the filing date. Further, the Examiner has made no showing that the posting was publically disseminated to anyone. On the contrary, web pages must be “browsed” — the antithesis of “disseminated.”

Accordingly, the www.xactware.com posting is not prior art within the meaning of 35 U.S.C. 102, and the Examiner has not made a *prima facie* showing of obviousness under 35 U.S.C. 103(a) with respect to any of the pending claims. The Applicants' respectfully request that the Examiner's rejections be withdrawn, and that the pending claims be allowed.

Respectfully submitted,

WILLIAM T. ROWSE et al.

By 
John S. Le Roy
Reg. No. 48,158
Attorney/Agent for Applicant

Date: Nov. 23, 2005

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351



The Shifting Dynamics of the Content Market
April 9-11, 2006
Carnegie Inn | Scottsdale, Arizona

[Table of Contents](#)[Previous Issues](#)[Subscribe Now!](#)

VOLUME 26 • NUMBER 2 • MARCH/APRIL 2002

• ON THE NET •

The Wayback Machine: The Web's Archive

by Greg R. Notess

Reference Librarian, Montana State University

Now that the Internet is established in the public information space, it has become a new publishing medium. The Web in particular has proved an incredible repository of all kinds of information content. But it has also proven to be a very changeable medium, noticeably lacking in permanence. Particularly during the past couple of years, as the number of new dot com failures has risen, previously existing Web sites have ceased operations and their information content has vanished into the Web's past.

With print publications, the libraries and archives of the world have made a major effort to collect and preserve print items. But the advent of the Web was so sudden and created an entirely new set of problems for cataloging, storage, and retrieval, that few libraries actively collected copies of Web pages. While the library profession worked diligently on finding solutions to the access side of the problems, Web pages were created, changed, and died, with no record of those pages being retained.

Fortunately, Brewster Kahle's Alexa Internet and its sister company, the Internet Archive, have done a huge amount of the collection work. Starting in 1996, the Internet Archive has been storing Web pages, including graphics files, from publicly accessible Web sites that Alexa has crawled. With the October 2001 launch of the Wayback Machine, this huge archive is now freely available to the Web public.

WHAT WAYBACK DOES

The Wayback Machine is a front end to the Internet Archive's collection of public Web pages. It includes more than 100 terabytes of data—a huge collection with huge storage requirements. The Wayback Machine provides access to this wealth of data by URLs. It is not text searchable—a user needs to know the exact URL of a particular Web page, or at least the Web site, to be able to enter the archive.

Upon entering an Internet address, the Wayback Machine presents a list of dates showing when that particular page has been archived. A check on the home page for the Library of Congress finds archived pages from November 9, 1996 through yesterday. There are far fewer pages in the 1996, 1997, 1998, and 1999 archives. In 2001, there was a copy from almost every other day.

Click on one of the displayed dates to see the archived page. The asterisk after some of the dates is used to designate when the Internet Archive detected a change in the page. So presumably, all those listings without an asterisk should be exactly the same as the first page before them that has an asterisk.

Note that the URL for the archived page begins with web.archive.org. Unlike the cached files at Google, the Wayback Machine also includes most image files in the archive. Thus, the images are not being drawn from a current server, but from the Internet Archive itself. This means that the archived page will display much more accurately how the page appeared on that particular date.

In addition, all the links on an archived page point not to the original linked location, but to other pages in the Internet Archive. So while the Wayback Machine is not searchable, it can be browsed. Find an archived page from 1997, click on any of the links on that page, and the Wayback Machine will take you to the closest (in terms of date) archive of the page available. In this way, a user can browse a Web site as it appeared within a certain time period.

The location of the Wayback Machine itself has shifted around among several URLs during its first few months. Both <http://web.archive.com> and <http://archive.alexa.com> worked in the past, but at this point, they all redirect to www.archive.org, the home of the Internet Archive itself.

WHY WAYBACK

There are many uses for the incredible archive from the Wayback Machine. At a very basic level, it is a great source to find the information on pages when the page or host itself is unavailable. When you come across a "404 not found" or similar message on the Web, just check on the Wayback Machine to find a copy of the page as it used to look. Google's cache used to be the only option for this function, but the cached pages are limited by the absence of any record of the date when they were cached. The Wayback Machine makes this so much easier by clearly identifying the date when the page was archived.

The historical implications of the Wayback Machine are immense. Historical researchers can now view significant portions of the Web as it existed at various times from 1996 to the present. The historical advantages go well beyond the pure historical research. Patent searchers can verify prior art. Business experts can look up failed companies' business plans. Employers can investigate job applicants' student Web pages. Sources lost because of complex URL shifting can be found by their old URL on the Wayback Machine.

The ability to view a range of versions of a particular page, and to browse the archived site itself, offers a range of uses. A new Web designer can look at previous incarnations of a site, even if the organization itself never archived the various versions. A new business can look at their competitors' early designs and avoid the same mistakes. And the researcher who is trying to track down the online resources from the bibliography of a four-year-old paper can find them in the archive, even if they have otherwise vanished from the current Web.

For institutions, the Internet Archive welcomes collaborative efforts to build special, theme-oriented collections. Already, there are three collections available: The September 11, 2001 collection, Web Pioneers, and Election 2000. As more special collections are created, they can be especially useful for more in-depth re- search on those topics.

ADVANCED FORM

Basic access to the archive is by a single URL, but the Wayback Machine also has an advanced search form. It is not linked from the front page, but is available as a link in small print at the top of the search form that appears with the results after a search has been entered. Look to the right of the "Take Me Back" button in the archived pages from the Library of Congress. It is also

directly available (<http://web.archive.org/collections/web/advanced.html>).

While there is still no textual search capability on the advanced search form, it does offer a range of options beyond the simple box on the home page. For example, the advanced form allows two kinds of URL Matching "Retrieve page that most closely matches search criteria" and "List all pages that match search criteria." The latter is the default on the simple form and brings up the list of date matches. The first option takes the user directly to the most recent copy of the archived page.

The advanced search form also gives options for limiting results to a specific range of dates. The individual archived pages have URLs so that they can be linked to directly. The advanced search page also explains the syntax. For example, the URL web.archive.org/20011230221317/http://www.site.net would be the www.site.net page archive on December 30, 2001 at 10:13p.m. and 17 seconds. In other words, the long list of numbers after the archive.org part represent the year, month, day, hour, minute, and second the page was archived in the form of YYYYMMDDhhmmss.

In addition to the scripted date limits available on the advanced form, an asterisk can be used as a truncation symbol within a URL as well. So, http://web.archive.org/200112*/http://www.site.net would retrieve a list of all the archived pages from December 2001. Leave off the asterisk and the Wayback Machine will automatically look for the page closest to the middle of the month. The truncation symbol can also be used to find all the pages from a site for a specific date. In other words, web.archive.org/1997*/http://www.site.net finds all the site URLs (pages and images) in the archive from 1997.

FILE FORMATS AND ALIASES

The advanced search form also points out that the Wayback Machine provides access to more than just Web pages. The File Types limit includes six formats: Images, Audio, Video, Binary, Text, and PDF. By choosing one of these file types and then only putting in the root URL (with a complete host name), the results will include all the file types of that format from that host in the archive. Each individual file type record has a unique URL, but if the searcher does not know the full URL, this limit helps to identify them. In addition, it can be used as a tool to count the number of a specific file type on a specific server.

The aliases are another nice feature on the advanced search. Many Web sites have multiple ways of writing a URL that will get to the exact same page, especially on the home page. The Aliases section of the advanced search gives three options. The default groups all host name aliases together, for the most comprehensive retrieval. However, a second option to "Show Aliases Separately" will give the exact matches for only the URL entered with a list of the other aliases while "Don't Show Aliases" will only give the exact matches.

LIMITATIONS

While recognizing the significant accomplishment of the Wayback Machine, it does have its limitations. Even with 100 terabytes of data, there is a great deal missing. The Internet Archive only includes a small amount of material from 1996, and the Web certainly pre-dates that. In addition, the older gopher content and other non-Web files are unavailable.

More significant are the orchestrated exclusions. Anyone can exclude their own pages by use of a

robots.txt file on their server. If the Internet Archive includes your Web pages and you want them excluded, just add a robots.txt file to exclude their crawler. The next time your page is crawled, all the old pages in the archive will be excluded as well. See www.archive.org/internet/remove.html for more details.

Unfortunately, far too many sites have had a robots.txt file excluding crawling or archiving. At least when a user requests a page that has been excluded due to a robots.txt file, the Wayback Machine gives an explanation as to why the page has been excluded and links to an archived copy of the site's robots.txt file.

The archiving process does have some problems. Most images are archived, but some still point to the original source and, thus, may end up as dead links or changed image files. Other images or objects on a Web page, especially at high traffic sites, may be linked to a network caching version, with a URL on an Akamai host, for example. Thus, some images on some pages will be missing.

Nor will the Wayback Machine always be available. After it first launched, a message often appeared stating that due to a "higher than expected number of requests," the Wayback Machine was down. At other times, you may run across a "This Internet Archive site is currently down for maintenance" message.

Given the huge size of the archive, another concern is the long-term financial viability of the Wayback Machine. Other than an Amazon button for donations, there are no ads on the site, nor does Alexa support it financially. According to Brewster Kahle, private fund raising, foundations, and grants currently support it. Kahle says that they "have enough to sustain the Wayback Machine, but that growth will be dependent upon financial support via joint projects."

Kahle should be lauded for trying to support the Wayback Machine more like a traditional library or archive as opposed to a typical commercial Web venture. The main page lists several donors including AT&T Research, Compaq, Prelinger Archives, QuantumDLT, and Xerox PARC. After all, Kahle hopes the Internet Archive can "build universal access to human knowledge. That's our goal in life."

It is a wonderful and worthy goal. And while the Wayback Machine has many limitations and excludes a huge amount of both online and print knowledge, it is certainly a major step forward in providing access to a large piece of that knowledge which is residing on the World Wide Web.



Greg R. Notess (greg@notess.com; www.notess.com/) is a reference librarian at Montana State University and founder of **SearchEngineShowdown.com**.

Comments? Email the editor at marydee@infotoday.com.

[\[Contents\]](#)

[\[ONLINE
Home\]](#)

[\[Subscribe\]](#) [\[Top\]](#)

[\[Information Today,
Inc.\]](#)